知识元理论研究述评

■ 索传军　戎军涛
中国人民大学信息资源管理学院　北京 100872

摘　要：[目的/意义]知识元作为知识的基本组分，是知识在微观领域的存在形态。对知识构成组分问题的研究体现了人类对知识认识的深化，是人类对知识认识的必然趋势。因此，对知识元及其运动规律的研究具有重要的理论价值。[方法/过程]首先从学科史角度对知识基本组分相关学说和知识元理论研究的现状进行梳理和分析；在此基础上归纳总结知识元理论研究中存在的问题；最后论述知识元理论研究的未来走向。[结果/结论]当前知识元理论研究存在以下问题：概念界定模糊、类型划分存在歧义、基于知识元的知识表达尚存在较大难度、相关研究尚未深入到知识的本质与运动规律。同时认为从多学科交叉视角开展知识元理论探析，将成为今后重要的趋势。如哲学与认知科学相结合的知识论视角、语言学与计算机科学相结合的知识发现与挖掘视角、知识工程与图书馆学相结合的知识表示研究视角、图书情报学科与出版学科相结合的知识组织研究视角等。

关键词：知识基本组分　知识元　知识元类型　知识元结构　知识元描述模型
分类号：G250.2
DOI：10.13266/j.issn.0252－3116.2021.11.014

1 引言

从学科发展史来看，各学科对知识的研究由来已久。从苏格拉底首先提出“什么是知识”的问题开始，知识就成为哲学领域的重要研究命题，关于知识的专门研究形成了哲学领域的“知识论”。政治学、经济学、管理学等其他学科也均有对知识的论述。从知识内容的生产、传播、交流、存储、管理与利用的全流程来看，图书情报学和出版学等学科当前仍然停留或局限于记录知识内容的文献，把记录知识的文献等同于知识，把文献管理与服务等同于知识管理与服务，影响了知识生产、传播、交流与利用效率。因此，整体上对知识本身的认识和探索比较浅显，尚没有深入到知识的内部，未能从微观上认识知识的构成与运动规律。

从宏观世界到微观世界的探索，是人类认识发展的基本方向。自然科学的发展历程表明，现代科学就是从打开微观世界的大门而开始形成的。对微观世界基本规律的掌握，是学科发展水平的重要标志，往往代表着学科的前沿和发展方向。知识元作为知识的基本组分，是知识在微观领域的存在形态。对知识构成组分问题的研究体现了人类对知识认识的深化，是人类对知识认识的必然趋势。因此，对知识元及其运动规律的研究具有重要的理论价值。

近年来，随着大数据、人工智能和知识服务业的快速发展，什么是知识，“知识”如何表示、组织、管理与服务等问题成为图书情报学、出版学、知识工程与人工智能等领域的研究热点。不同学科从不同的视角，对知识构成的基本组分（如知识元）进行了探究，初步形成了一定的共识，但对其内涵与外延、构成与特征、分类与重组等重要理论问题均未形成较为统一的科学认识，人们对知识元的认知，尚停留在“意会”阶段。因此，对知识元相关理论研究进行梳理，厘清研究现状和存在的问题，明确未来研究路径，具有重要的现实意义和价值。

2 知识基本组分问题的提出

2.1 知识基本组分学说图谱

知识组成单位究竟是什么，如何界定，知识组分间有何关系，围绕这些问题国内外不同学科的学者基于各自的学科视角先后提出了知识单元、知识因子、知识基因、知识元等概念。通过对文献内容和引用关系的分析，初步形成了知识基本组分学说的发展脉络图谱，

作者简介：索传军（ORCID：0000－0002－7416－1531），教授，博士，博士生导师，E-mail：suocj@ruc.edu.cn；戎军涛（ORCID：0000－0002－6242－0989），副研究馆员，博士研究生。
收稿日期：2020－11－30　修回日期：2021－03－15　本文起止页码：133－142　本文责任编辑：徐健

见图 1。其中,节点表示某一时期知识元基本组分学说的代表性观点,其包含 3 个属性,即观点提出的学者、观点提出的时间与包含此观点的代表性文献;边描述的是节点之间的关系,在该图谱中包含继承、顺承、发展和相关 4 类关系。

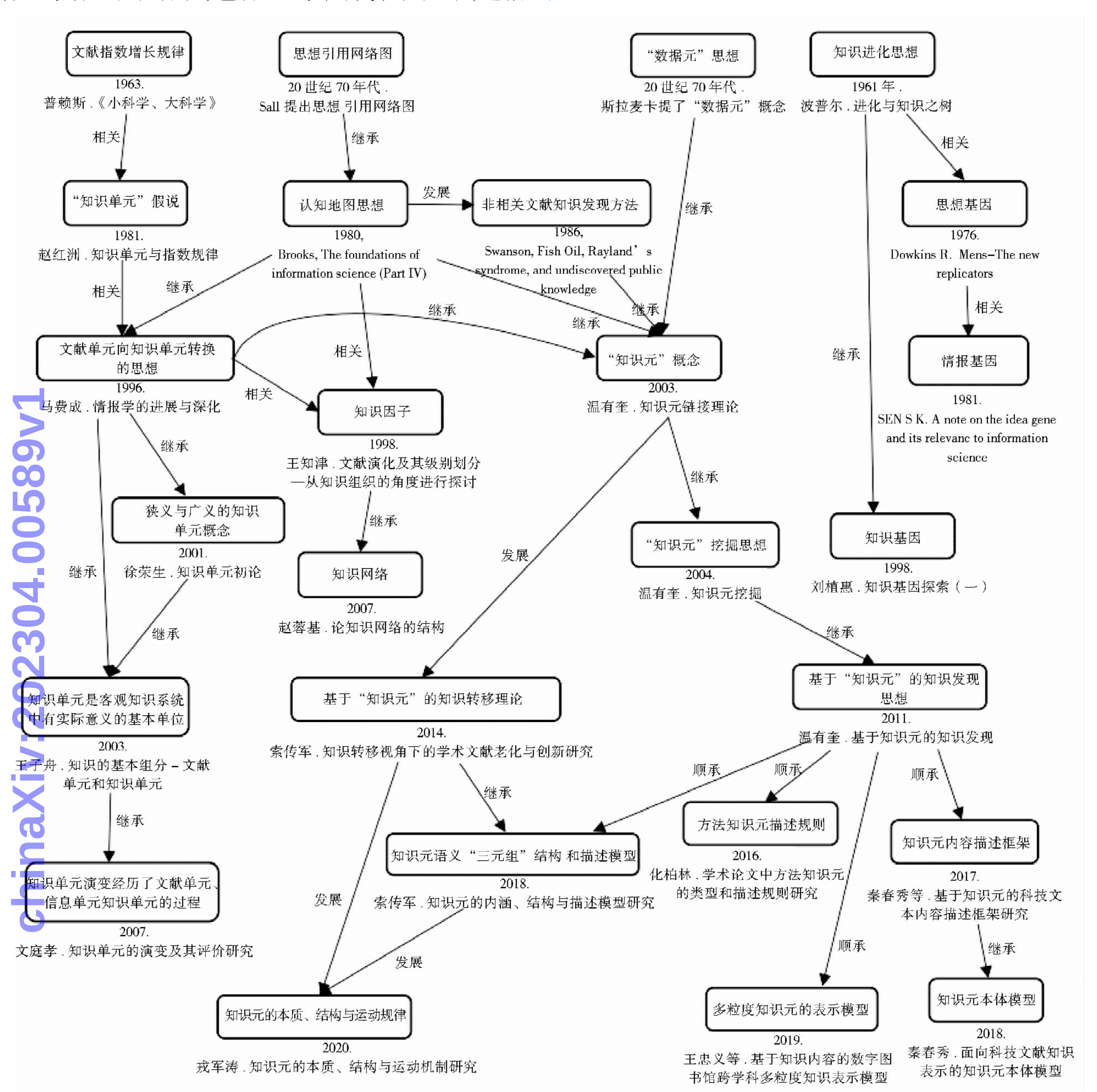


图 1　知识基本组分不同学说的脉络发展图谱

在知识基本组分的学说发展图谱中,布鲁克斯的认知地图思想和波普尔的知识进化论具有重要的理论影响。在认知地图思想影响下,分别形成了知识单元、知识因子与知识元概念,在波普尔的知识进化思想影响下,产生了思想基因、情报基因和知识基因理论。

2.1.1　认知地图学说

亨利・斯摩尔(H. Small)[1]提出用简单的概念陈述来表达论文的思想,进而构建学术论文的思想引用网络图。1980 年,布鲁克斯(B. C. Brookes)[1]受到斯摩尔的引用模式的启发,提出构建学科认知地图的设想。受普赖斯指数规律影响,赵红洲与蒋国华[2]从知识计量角度提出了"知识单元"的假说,然而这一思想并未对图书情报领域产生实际影响。1986 年,D. R. Swanson[3]提出了非相关文献的知识发现方法,为认知地图思想提供了技术路线。

受认知地图学说影响,武汉大学著名情报学家马费成[4]提出知识表达和组织必须从物理层次的文献单元向认识层次的知识单元或情报单元转换。该思想对

索传军，戎军涛.知识元理论研究述评[J].图书情报工作,2021,65(11):133-142.

我国知识组织领域具有重要影响。南开大学王知津[5]认为知识单元无非就是概念。针对知识单元的概念，徐荣生[6]提出知识单元有广义和狭义之分;北京大学王子舟等[7]从图书情报工作的实践出发，定义知识单元是客观知识系统中有实际意义的基本单位;文庭孝[8]进一步指出知识元的不同排列组合构成不同的知识单元，不同的知识单元按照不同逻辑关系组成不同的知识元链接，这一观点揭示了知识元、知识单元与知识结构的关系。

20世纪70年代，美国情报学家弗拉基米尔·斯拉麦卡(V. Slamecka)教授在华讲学时提出数据元的概念[9]。受到布鲁克斯和斯拉麦卡的影响，西安电子科技大学温有奎等[10]较早提出了知识元概念，其强调知识元是一个由若干个必要的要素组成的实体，由向导信息和知识组成[11]。

2.1.2 知识基因学说

20世纪50年代，英国著名哲学家波普尔[12]在深入研究了达尔文的进化论后，发现知识的发展与生产进化有着惊人的相似之处，并于1961年发表重要论文《进化与知识之树》，提出人类知识是长期进化的产物，这一点符合达尔文的进化论模式，认为知识进化中蕴含着"基因"。1976年，道金斯(R. Dowkins)[13]出版专著*Mems-The new replicators*，提出了"思想基因"的理论。1981年，印度情报学家斯·科·森(S. K. Sen)[14]进一步提出"情报基因"的概念。

在波普尔的知识进化思想和认知地图思想影响下，王知津等[15]提出知识因子概念，其认为可将知识视为一种网状结构，这种特定意义上的知识就是由众多结点(即知识因子)和结点间联系(即知识关联)两个要素组成的。赵蓉英[16]继承了知识因子说，在此基础上进一步阐明了知识网络的结构和功能。知识基因理论源于生物进化论，我国学者刘植惠受波普尔知识进化理论启发，提出了知识基因理论，主要包括知识基因的含义、结构、内涵、形态、特征及属性等。知识基因理论认为知识基因是知识继承与发展的最小单元，具有稳定性、统摄性、遗传与变异性以及定向控制知识走向能力4个特征[17]。知识基因可以构成粒度更大、层次更高的知识DNA(主要指科学定律)[18]、知识细胞(表现为科学理论、学说、主义等)[19]和知识器官(表现为知识学科)[20]。

2.2 知识基本构成组分各学说比较分析

通过图1可以看出知识单元、知识元、知识因子均是受认知地图思想影响而产生的知识组分思想，而知识基因则是波普尔知识进化思想的进一步发展和深化。

2.2.1 知识单元与知识元

从目前对于知识单元的认识上来看，知识单元与知识元是两个截然不同的概念。知识元强调独立性、完整性、单一性，是不可再分的、最小粒度的知识单元。知识元的排列组合构成了不同粒度的知识单元。知识单元强调完整性、独立性、层次性、多粒度性，知识单元有大有小，层次有高有低，只要具有知识完整性、独立性，就可以成为一个知识单元。因此，知识单元指一个知识单位，并非专指知识的最小组分。知识单元是知识元的上位概念，涵盖了知识元的内涵，是知识元的重要相关概念。厘清知识单元和知识元的关系，是知识元的基础理论研究必须解决的任务。

2.2.2 知识因子与知识基因

知识因子说从知识网络角度探讨了知识的基本组分问题。其认为知识因子是组成知识的最细微的成分，知识是由知识因子为节点构成的知识网络。知识节点是知识网络的节点，知识节点并不强调知识粒度的大小，任何粒度的知识单元均可充当知识节点。知识节点并非不可分割，可以由更小的知识节点组成，这体现了知识节点的网络嵌套性，即知识网络可以充当知识节点，构成更大的知识网络。知识节点通过关联构成知识链接，知识链接关联构成立体的知识网络，这对于知识元链接、知识元网络构建具有启发意义。

知识基因说将生命体与知识体系进行类比，发现两者具有极大相似性。知识基因说所提出的系列概念、原理，能够解释知识体系的不同组成单元，从微观的知识基因遗传和变异视角解释了知识的传承、走向和创新问题，对知识元的理论研究具有极大的启发意义，开拓了知识元研究的视角，引发了学术界广泛的关注。但是，知识基因如何组成更大一级知识有机体，知识基因的结构，知识基因如何表达，知识基因说并未对这些基本问题给出结论。此外，知识基因解释了纵向的知识结构以及知识的发展创新问题，但还未能解释横向上的知识引用引起的知识转移、老化问题。知识基因学说从仿生学视角对知识创造进行了研究，特别是知识基因突变思想，揭示了知识演化的内部动因，具有较大的学术价值，但该理论还有待进一步实证。知识基因理论应该从知识进化系统论的观点深入探索知识本质，应该完善和发展知识测度方法，实现知识程序化和工程化，进一步扩大知识基因的应用领域[21]。

总的来说，相较于知识单元、知识元、知识基因、知

识节点、知识因子等概念，知识元概念一经提出就引起了学术界的关注，并且得到了多数学者的赞同和支持。特别是知识工程领域对知识元的探索和研究较多，一些数据库服务商纷纷开展知识元的研究与实践，出版业也相继提出纳米出版物和微出版物等。但是不同学科的不同学者对知识元概念使用还较为混乱，因此亟需开展知识元的相关理论与实践研究，构建基于知识元的知识表达理论体系，提升知识管理与服务效率和效果。

3 知识元理论研究现状

自从知识基本组分的几种学说提出以来，形成了多学科的研究热点，相较于其他几种学说，知识元学说的研究相对较为深入和系统。近20年来，不同学科领域学者从理论、技术和应用层面开展了一系列研究，形成了以“知识元”为研究对象的学术领域。

知识元的理论研究任务是描述知识元的本质、结构以及存在状态，其目的是揭示知识元运动的一般机理，充分发挥理论的认知、解释、预测和指导功能，从根本上推动知识元应用。知识元的技术是指在知识元获取、标注、存储、管理和服务等过程中所采用的具体方法。知识元的应用是指利用相关技术解决知识活动中的各类问题，相关应用领域包括语义出版、科学计量、知识管理、数字人文和教育教学领域等。

从当前研究现状来看，对知识元的认识虽然初步达成了一定的共识，但对知识元运动规律还缺乏系统深入的研究。例如，知识元是如何发生转移、聚合和融合的，在知识的生产和创新过程中不同类型的知识元其作用和运动机理是什么。理论方面，目前知识元的相关研究还没有触及知识元的本质；实践方面，多是采用知识工程的技术路线，还没有大规模的、成熟的知识元应用范例。知识元理论研究的困境，在一定程度上制约了以知识元为基础的知识组织和服务实践的开展。因此，关于知识元理论研究将成为突破知识元技术与应用研究的关键问题。

3.1 知识元概念界定

知识元概念是知识组分理论研究的逻辑起点。多数学者从概念入手，对知识元进行初步研究。知识元概念的定义一方面要求有明确的外延边界，另一方面要求有深刻的理论内涵。当前，关于知识元概念的研究，主流观点可分为两种模式，即内容界定模式与外延界定模式。详情如表1所示：

表1 知识元概念界定模式

界定模式	内容描述	代表人物
内容界定模式	知识元的概念内涵界定包含知识单位、知识结构、知识片段、最小颗粒度、独立性、完备性等元素	温有奎等[22]、赵蓉英[16]、索传军等[23]
外延界定模式	指“知识元”这一概念所概括的思维对象的数量或者范围，是适用于概念的一切对象，即知识元概念的适用范围。例如一段文字、一幅图表、一个公式、一章或一节、一段动画、一个程序等	姜永常等[24]、文庭孝[8]

3.2 知识元分类

对知识元进行分类，是进一步认识、区分知识元的一种逻辑方法。不同类型的知识元具有不同的属性、结构、特征，因此知识元分类是知识元抽取、标引、挖掘、利用的前提和基础。但是，由于人们对知识元内涵认识还未统一，导致知识元的分类也呈现出多样性，其分类思想主要有认知分类思想、内容分类思想与功能分类思想，如表2所示：

表2 知识元类型划分

分类思想	内容描述	代表人物
认知分类	利用认知心理学的知识分类思想对知识元进行分类，把知识元分为描述型和过程型两大类	温有奎等[25]、赵蓉英等[26]
内容分类	主要根据知识元所表达的语义内容对其进行分类，是一种语义分类思想。一般分为概念型、原理型、方法型、事实型、陈述型、数值型、模型型等	张静等[27]、原小玲[28]、毕崇武等[29]、高国伟等[30]、秦春秀等[31]
功能分类	根据知识元在文本中的语用功能进行分类，是一种语用分类思想。例如常识知识元、引证知识元和创新知识元。	廖开际等[32]、索传军等[23]

3.3 知识元的知识表达模型

知识元的形态通常表现为一段文本片段或一组语句。知识元的知识表达是指对知识元的形式化描述，以便于机器处理和理解，即将非结构化的知识转化为结构化知识的一种语义描述方法。根据文本内容的形式化程度，知识元知识表达模型可分为词组模式、语句模式和三元组模式，见表3。

3.4 知识元的描述模型

知识元的描述模型分为线性描述结构和分层描述结构。线性描述结构简单、方便，便于机器识别理解，是表达知识元属性的重要方法，如表4所示。但是随着知识元属性的增多，导致了元数据结构维度增高、描述增多、表达冗长和推理不完备等问题，增加了计算机对语义内容精确识别和理解难度。知识元描述的分层表示结构使得知识元的属性模型表达更加清晰，语义

表3　知识元的知识表达模型

表达模型	内容描述	代表人物
词组模式	利用词组模式进行知识元的内容表示	Z. Chen[33]、M. Weeber等[34]、Y. Illhoi[35]、Y. Tsuruoka等[36]、M. Yetisgen-Yildiz等[37]、T. Cohen等[38]、周宁等[39]、文庭孝等[40]
特征语句模式	把主题句作为知识元内容的描述	温有奎等[41]、郑彦宁等[42]、郑梦悦等[43]
三元组模式	知识元内部的文本内容要表示成计算机可理解的模式，通常借鉴知识工程领域的语义三元组形式，即利用“主谓宾”三元组来表示一段语句的内容	温有奎等[44]、白如江等[45]、索传军等[23]

化更明显，解决了由于属性增多导致描述结构不便于机器理解的困境，使知识元属性结构表达更加清晰、语义化效果更明显。

表4　知识元的描述模型

结构模型	内容描述	代表人物
线性描述结构模型	“N元组”一般体现为线性描述结构的形式，将有关知识元的属性元素进行罗列描述，形成线性结构。例如四元组、六元组、七元组	温有奎等[10]、马倩倩[46]、刘淼等[47]、余洋[48]、钱飞[49]、李珊珊等[50]、许春漫[51]
分层描述结构模型	知识元描述元素都可以归类为内容、属性（内容特征和外部特征）和关系等类别	J. H. Zou等[52]、谢庆球等[53]、毕崇武等[29]、索传军等[23]、王忠义[54]

4　知识元研究存在的问题

从目前研究来看，越来越多的学者认为，知识组织的控制单位应该深入到文献内容的细粒度知识元。但关于知识元的基本认识还很模糊，缺乏逻辑思辨和抽象认识，缺乏对知识元的本质及其运动规律的深入系统探索。理论研究呈现出经验性描述的特征，制约了以知识元为基础的应用研究与实践的深度。

4.1　知识元的概念界定模糊

由于知识元所涉及问题的复杂性，当前人们对知识元的认知还仅仅停留在观点、结论、概念、公式等表层形态层面。关于知识元概念的内涵与外延仍然较为模糊，也无法做到对知识元类型的清晰界定。

4.1.1　对知识元的本质认识不明确

从形式上看，知识元通常表现为一段文本内容。因此，多数学者从文本内容视角对知识元的外延进行表面的、经验式的描述，对知识元的内涵揭示不够全面，缺乏对知识元特征和属性的描述，因此“知识元”仍是一个不确定的概念。

4.1.2　知识元的粒度大小（或形态）不明确

相对完整的表达一条知识是知识元的基本内涵要素，这要求知识元必须是一条完整的知识表达。然而知识的完整性并不代表知识粒度的最小化，这需要清楚地界定知识元的完整性与粒度大小的关系。

4.1.3　知识元内容与语言符号相混淆

从语言学角度来看，知识元是一段语言符号片段；从语义学角度来看，知识元是一段完整的语义概念表达。如何界定二者的关系，是知识元概念研究必须解决的理论问题。

4.2　知识元分类存在歧义

（1）认知分类思想。将知识元分成描述型和过程型两大类，这种划分易于实施，但过于集中，不能充分描述不同情况和不同过程中知识元的类型。

（2）内容分类思想。相对能够充分地刻画知识元的类型，是目前知识元分类的主流思想，但是也存在一些弊端：①知识元的分类标准不统一，不同类别之间存在交叉、蕴含关系，不易区分；②知识元类型划分过于分散，容易引起语义歧义，不利于计算机识别和抽取；③知识元划分依据主要是专业领域，缺乏通用的分类思想与方法；④各类型知识元之间缺少一定的逻辑支持关系。

（3）功能分类思想。这是一种比较均衡、综合的知识元分类方法，具有标准统一、类目适当的优点，可以克服上述知识元类型划分的弊端，但是该思想目前还未占主流。

概括地说，知识的存在分为显性知识和隐性知识。知识元作为知识的组成要素，显然也具有这两种存在状态，即显性知识元和隐性知识元。从当前的知识元类型来看，并未涉及知识元的隐性状态，因此概念内涵组成和外延的界定都缺乏必要的的隐性知识元要素。

4.3　知识元描述存在难度

当前关于知识元描述基本上采用的是词组模式、特征语句模式和三元组模式。

（1）词组模式。利用词组模式来表示知识元的文本内容，具有简单、方便、高效、利于计算机抽取和识别等优点，回避了中文自然语言切分和理解的困境。但是众多语义仅使用单词、双字词和三字词是无法概括和准确表达的，尚没有解决文献知识片段的语义关系问题。

（2）特征语句模式。利用句子表征知识元的内容，虽然克服了用特征词进行知识表示的弊端，保留了全部的语义，但是对计算机来说仍然是一段不能理解的文本字符串，没有解决语义结构化的问题。因此，这种模式虽然易于实现，但也只是一种文本知识表示的过渡模式，最终文本要表示成计算机可以理解的结构化数据模式。

（3）三元组模式。语义三元组是最简单的一组语义表达的数据模型，其本质是一种实体关系模型，需要自然语言处理技术获取实体、关系、属性及属性值等语义角色和语义依存关系，经过形式化语言如 RDF 描述后可以被计算机理解。不同的语义三元组之间可以通过语义链接相关联，构成更广泛的语义网络，因此可以描述更加复杂的文本内容。

尽管理论上可以将知识元内容表达为机器可理解的“三元组”模式，但是由于知识元内容承载和表达了丰富的语义信息，含有大量丰富的词法特征、句法依存和语义依存关系，不同知识元之间也存在复杂的逻辑组合关系，且并非所有的文本都适合用“三元组”表达，所以“三元组”这一构建过程非常困难。

4.4 知识元研究尚未深入到知识的本质、结构与运动规律

事物的本质、结构是运动规律的基础，规定、制约着运动的机制、程度、方向与范围。事物的运动规律恰恰反映了事物的本质和结构状态，是事物本质的、内在联系的表现。理论研究的任务就是揭示事物的本质、结构、存在状态以及运动规律。其目的是，阐释事物存在的现象和原因，系统揭示事物存在和发展的一般机理，充分发挥理论的认知、描述、解释、预测和指导功能，从根本上加速应用的变革与实践模式。

（1）知识元研究有待从“语言符号层向语义内容层”发展。当前知识元理论研究的前提和假设是，将知识元视为一段内容文本片段（属于语言符号）。通常情况下，知识元虽然表现为一段文本，但是知识元的本质并不是文本，应该从更深刻的语义、概念层次进行理论抽象。相关研究尚未涉及知识元本质、结构与运动规律等基础理论问题，无法发挥理论认知、解释功能，制约了以知识元为基础的知识组织和知识服务实践与应用的开展。

（2）知识元研究有待从“知识元自身向运动机理”发展。当前知识元研究，较多的是知识元的概念、类型、特征、属性以及知识表达模型和描述模型，侧重知识表示研究，较多呈现出经验描述与总结的特点，缺乏知识元转移与老化、聚合与融合，缺乏知识元在知识创新中的作用与机理等研究。

5 知识元研究展望

如前所述，知识元理论研究存在诸多问题，一是关于知识元的本质、特征、属性、粒度、结构、分类与功能等基本问题的科学认识；二是知识元运动规律的认识。基于上述问题的探索，构成了今后知识元理论研究的重点发展方向。

5.1 研究视角

知识元问题的研究，涉及哲学、图书情报学、知识工程、出版学、教育学、语言学等学科，具有多学科研究视角。可以借鉴不同学科的理论和方法，进行跨学科交叉研究，实现认识上的统一，是未来知识元研究的方向。

5.1.1 哲学与认知科学相结合的知识论视角

哲学与认知科学是研究知识的基础学科，围绕知识本身的基本认知问题就形成了知识论的研究视角。研究内容包括，知识的概念、本质与功能、知识结构与形态、知识的形成、转移与老化等内容。知识论视角强调理论的内在逻辑性、完整性与抽象性，通常采用归纳演绎、逻辑抽象等理性主义路线。知识论视角的研究有助于厘清知识元的概念、类型、本质、结构、存在状态以及运动规律等根本问题，为知识元理论构建提供坚实的基础。

5.1.2 语言学与计算机科学相结合的知识发现与挖掘视角

人们通常以语言为载体进行知识交流，因而从形式上看，知识元通常表现为一段文本，所以多数学者对知识元的概念、类型、模型等理论研究基本上以语义概念、符号表达为理论基础，呈现出语言学、语义学视角。在语法层面上，知识表现为一段文本，基本组成结构一般是语词、语句、段落、节、篇、章；在语义层面，知识的最小构成单位是句子，词是概念的符号，一般不能表达一条完整的知识。该视角以格语法、框架语义学、语义角色标注、深层语义分析等理论为基础进行语句的句法分析和语义依存分析，为知识元的形式化表达提供了理论支持。

5.1.3 知识工程与图书馆学相结合的知识表示研究视角

当今万维网的飞速发展,知识正以前所未有的速度快速增长。知识的表现形态也已突破人们已有的认知,蕴含在各类媒体的大数据之中。知识图谱是大数据环境下一种基于图的知识表示的方式,是大规模语义网,是知识库,其核心是各类实体及其关系构成的语义网络。知识工程领域通常利用计算机自然语言处理技术实现对文本中包含的实体进行识别、抽取与表示。图书馆学的优势是利用各类词表等构建描述知识的本体,也就是知识图谱的数据模式。知识图谱研究视角需要利用知识工程的技术方法将各种数据中表征知识元内容的实体进行识别、理解、筛选、归纳,图书馆学需要利用本体构建技术与方法,构建知识图谱模式,并存储为知识元库。

5.1.4 图书情报学科与出版学科相结合的知识组织研究视角

知识组织视角是图书情报和出版领域采取的研究模式。图书情报领域关注知识元的知识表示、描述模型、标引规范等规范控制理论。出版领域关注单篇文献内知识基本组分(内容单元或数字对象)的组织与呈现方法与方式的研究,特别是不同媒体、不同类型数字对象的组织。例如,纳米出版物对文献中断言、数据等的描述。基于知识元的知识组织理论将在语义出版领域取得应用。对文献所包含知识元进行描述、组织、测度和分析,将是图书情报与出版领域今后重要的研究视角。

5.2 有待重点研究问题

基于上述对知识元研究现状和存在问题的分析,未来需要基于多学科合作交叉视角对以下问题进行研究。

5.2.1 关于知识元的概念、类型与模型等问题系统的研究

(1)知识元概念。知识元的概念、类型、本质等问题属于知识元的基础理论问题,这是整个知识元理论体系的根基,首先需要从哲学和认知科学视角界定什么是知识,明确对知识的认知,以此为基础来构建抽象的知识元逻辑概念,明晰知识元的类型,逐步深化对知识元的本质的认知。知识元的本质是知识元自身内部稳定的、一贯的、深刻的属性,决定了知识元与非知识元的边界。只有确定了明晰的本质,才能界定知识元的概念。

(2)知识元特征。从现有关于知识元概念的论述中可以得知,知识元具有独立性、完整性、稳定性、唯一性、不可再分性等特点。通过内容分析,可以发现知识元首先是一条完整的知识,语义表达的完备性是其本质特征。其次,知识元具有内容上的相对独立性,知识元可以被独立进行语义描述,可以被重新融合,形成新的知识。最后,知识元具有语义与逻辑上的不可再分性。因此,知识元应该是知识结构中相对独立的、具有完备知识表达、最小粒度的语义单元。

(4)知识元形态。关于知识元的粒度,也就是知识元是否可拆分这一关键问题,目前还存在较大争议。笔者认为知识元从语义上是不可拆分的,若拆分其语义会发生改变,就不是该知识元了。当前研究还未能从知识元本身的客体视角对知识元的内部结构进行理论抽象。

(5)知识元类型。知识元的类型也是一个有待研究的重要问题。知识元是知识的基本组分,其分类是否受知识类型(或分类)的制约;不同类型的知识元,其表现形态与结构、功能与运动方式等方面是否存在不同等都是有待研究的重要问题。

5.2.2 基于知识元的知识表示问题研究

基于知识元的知识表示是对知识元及其关系的形式化描述与表征的过程,其结果是形成计算机可识别、可理解、可操作的结构化知识,这是知识组织与服务从文献层面深入到知识层面的关键环节。基于知识元的知识表示一般包括基于知识元的文献本体构建和知识元标注两个关键过程。

一篇结构良好的学术文献由若干知识元组成,但并不是知识元的简单堆砌,而是一组知识元的逻辑组合[55]。基于知识元的文献本体是用来描述文献内容知识元的语义数据模型,是利用本体技术将文献文本描述成知识元概念集合的形式化的规范说明,是控制文献内容知识元进行逻辑组合的数据结构,是基于知识元进行知识表示的核心模型。这是语义网环境下数据出版的重要研究内容。

5.2.3 关于知识元的运动规律研究

知识元是构成知识的基本组分。显然,知识元的运动规律与知识的产生和老化有直接关系。因此,对知识元运动规律的研究,对于知识的生产、传播、交流和管理等具有重要的意义。知识元的运动主要指知识

元的相互作用过程。其运动规律包括知识元的自我演化和不同知识元的融合两个方面。知识元自我演化包括知识元自身的形成、发展、转移、流动、转化、老化、衰减等;知识元的融合包括不同知识元的作用产生新知识、知识的扩散转移、知识的老化与消亡等。对知识元运动规律的研究,不仅有助于提高对知识元的认识,而且对知识运动与管理都有极为重要的价值。

(1)知识元的计量。对知识元的计量是从数量上统计分析知识元的运动规律。知识元的计量主要研究包括:知识元的单位测度研究、知识元关联度测量、相似度计算、知识元生产与分布规律、知识元指数增长规律、知识元的老化与转移规律、知识元的演化规律等。

(2)基于知识元的知识创新与生产。基于认知科学视角的知识元的融合创新研究,侧重探索知识元融合和知识元形成的关系,观察知识元产生过程中知识元内部结构的排列组合变化方式,从知识元的内部结构层面揭示具有创新价值的知识元的产生机制,有助于揭示认知科学领域中创新性知识的产生机理,进一步为人工智能领域提供人机理解基础。

6 结语

从科学技术发展史来看,有关"知识"问题的研究由来已久,但是关于知识及其基本构成组分的理论研究尚不成熟。伴随着知识经济的快速发展,数据、信息、知识成为社会发展的重要资源,基于"知识基本组分"的知识运动规律的相关理论研究成为图书情报学和知识工程等多学科关注的热点问题。不同学科围绕知识元的获取、描述、表达、存储、管理等问题进行了研究,并在理论层面对知识元形成了一定的认知。从目前研究来看,越来越多的学者认为,知识组织与控制(或管理)单位应深入到文献内的细粒度知识元。但关于知识元的基本认识(内涵与外延,形式与语义等)还很模糊。一是关于知识元的理论研究呈现出经验性描述的特点,尚未形成关于知识元的本质、特征、属性、粒度、结构、功能等基本问题的深层次的理性认识;二是理论研究没有触及知识元的运动规律,例如知识元的产生、流通、分布、重组、分类、计量等问题,导致基于知识元的知识组织与服务研究陷入瓶颈期,无法取得突破性进展。因此,知识元的研究必须要承担起艰巨的理论构建任务,揭示知识元的本质及运动规律,构建完善的理论体系。

笔者认为知识元研究涉及众多学科,形成了多学科的交叉视角,可以从哲学与认知科学相结合的知识论视角、语言学与计算机科学相结合的知识发现与挖掘视角、知识工程与图书馆学相结合的知识表示研究视角、图书馆学与出版学科相结合的知识组织研究视角进行多学科交叉合作,开展知识元的概念、类型、描述模型、运动规律、知识表示问题等研究将成为今后重要的趋势。

参考文献:

[1] BROOKES B C. The foundations of information science (Part IV). information science: the changing paradigm[J]. Journal of information science, 1981, 3(1): 3-12.
[2] 赵红洲,蒋国华. 知识单元与指数规律[J]. 科学学与科学技术管理, 1984(9): 39-41.
[3] SWANSON D R. Fish oil, Rayland's syndrome, and undiscovered public knowledge[J]. Perspectives in biology and medicine, 1986, 30(1): 7-18.
[4] 马费成. 情报学的进展与深化[J]. 情报学报, 1996, 15(5): 338-344.
[5] 王知津. 知识组织的研究范围及发展策略[J]. 中国图书馆学报, 1998(4): 3-8.
[6] 徐荣生. 知识单元初论[J]. 图书馆杂志, 2001, 20(7): 2-5.
[7] 王子舟, 王碧滢. 知识的基本组分——文献单元和知识单元[J]. 中国图书馆学报, 2003(1): 5-11.
[8] 文庭孝. 知识单元的演变及其评价研究[J]. 图书情报工作, 2007, 51(10): 72-76.
[9] 徐如镜. 开发知识资源 发展知识产业 服务知识经济[J]. 现代图书情报技术, 2002(S1): 4-6.
[10] 温有奎, 徐国华. 知识元链接理论[J]. 情报学报, 2003, 22(6): 665-670.
[11] 温有奎. 基于"知识元"的知识组织与检索[J]. 计算机工程与应用, 2005(1): 55-57, 91.
[12] POPPER K R. Objective knowledge: an evolutionary approach [M]. Oxford: Claredon Press, 1979: 256-284.
[13] DOWKINS R. Mems-the new replicators[M]. Oxford: Oxford university press, 1976: 25-28.
[14] SEN S K. A note on the idea gene and its relevance to information science[J]. Annals of library science and documentation, 1981, 28(1/4): 97-102.
[15] 王知津, 王乐. 文献演化及其级别划分-从知识组织的角度进行探讨[J]. 图书情报工作, 1998, 42(1): 4-7.
[16] 赵蓉英. 论知识网络的结构[J]. 图书情报工作, 2007, 51(9): 6-10.
[17] 刘植惠. 知识基因探索(一)[J]. 情报理论与实践, 1998, 21(1): 62-64.
[18] 刘植惠. 知识基因探索(七)[J]. 情报理论与实践, 1999, 22

(1):61－64.

[19] 刘植惠.知识基因探索(八)[J].情报理论与实践,1999,22(2):141－144.

[20] 刘植惠.知识基因探索(九)[J].情报理论与实践,1999,22(3):220－224.

[21] 刘植惠.知识基因理论的由来、基本内容及发展[J].情报理论与实践,1998,21(2):71－76.

[22] 温有奎,焦玉英.基于知识元的知识发现[M].西安:西安电子科技大学出版社,2011.

[23] 索传军,盖双双.知识元的内涵、结构与描述模型研究[J].中国图书馆学报,2018,44(4):54－72.

[24] 姜永常,杨宏岩,张丽波.基于知识元的知识组织及其系统服务功能研究[J].情报理论与实践,2007,30(1):37－40.

[25] 温有奎,温浩,徐端颐,等.基于知识元的文本知识标引[J].情报学报,2006,25(3):282－288.

[26] 赵蓉英,张心源.基于知识元抽取的中文智库成果描述规则研究[J].图书与情报,2017(1):119－127.

[27] 张静,刘延伸,卫金磊.论中小学多媒体知识元库的建设[J].现代教育技术,2005(5):68－71.

[28] 原小玲.基于知识元的知识标引[J].图书馆学研究,2007(6):45－47.

[29] 毕崇武,王忠义,宋红文.基于知识元的数字图书馆多粒度集成知识服务研究[J].图书情报工作,2017,61(4):115－122.

[30] 高国伟,王亚杰,李佳卉,等.基于知识元的知识库架构模型研究[J].情报科学,2016,34(3):37－41.

[31] 秦春秀,刘杰,刘怀亮,等.基于知识元的科技文本内容描述框架研究[J].图书情报工作,2017,61(10):116－124.

[32] 廖开际,熊会会,叶东海.基于知识元理论的应急文档结构化建模[J].计算机应用研究,2011,28(1):175－178.

[33] CHEN Z. Let documents talk to each other: a computer mod for connection of short documents[J]. Journal of documentation,1993,49(1):44－54.

[34] WEEBER M,KLEIN H, DE JONG-VAN DEN BERG LTW,et al. Using concepts in literature-based discovery: simulating Swanson's Raynaud-Fish oil and Migraine-Magnesium discoveries[J]. Journal of the American Society for Information Science and Technology,2001,52(7):548－557.

[35] ILLHOI Y. Semantic text mining and its application in biomedical domain[D]. Philadelphia: Drexel University,2006.

[36] TSURUOKA Y,TSUJII J,ANANIADOU S. Facta: a text search engine for finding associated biomedical concepts Bioinformatics[J]. Bioinformatics,2008,24(21):2559－2560.

[37] YETISGEN-YILDIZ M,PRATT W. A new evaluation methodology for literature-based discovery systems[J]. Journal of biomedical informatics,2009,42(4):633－643.

[38] COHEN T,SCHVANEVELDT R,WIDDOWS D. Reflective random indexing and indirect inference : a scalable method for discovery of implicit connections[J]. Journal of biomedical informatics,2010,43(2):240－256.

[39] 周宁,余肖生,刘伟,等.基于XML平台的知识元的知识元表示与抽取研究[J].中国图书馆学报,2006(3):41－45.

[40] 文庭孝,侯经川,龚蛟腾,等.中文文本知识元的构建及其现实意义[J].中国图书馆学报,2007(6):91－95.

[41] 温有奎,温浩,徐端颐,等.基于创新点的知识元挖掘[J].情报学报,2005,24(6):663－668.

[42] 郑彦宁,化柏林.句子级知识抽取在情报学中的应用分析[J].情报理论与实践,2011,34(12):1－4.

[43] 郑梦悦,秦春秀,马续补.面向中文科技文献非结构化摘要的知识元表示与抽取研究——基于知识元本体理论[J].情报理论与实践,2020,43(2):157－163.

[44] 温有奎,焦玉英.知识元语义链接模型研究[J].图书情报工作,2010,54(12):27－31.

[45] 白如江,周彦廷,王效岳,等.科学事件知识图谱构建研究[J].情报理论与实践,2020,43(9):107－114,124.

[46] 马倩倩.基于关联数据的知识元链接图式存储模型研究[D].武汉:华中师范大学,2014.

[47] 刘淼,王宇.基于主题句的期刊文献知识元库构建[J].情报杂志,2012,31(11):145－149.

[48] 余洋.组织管理中的知识超网络研究[D].大连:大连理工大学,2009.

[49] 钱飞.中文文献知识元描述研究[D].北京:中国人民大学,2014.

[50] 李珊珊,王宇.基于HNC理论的期刊文献知识元检索[J].情报杂志,2013,32(9):190－194.

[51] 许春漫.泛在知识环境下知识元的构建与检索[J].情报理论与实践,2014,37(2):107－111.

[52] ZOU J H,LIU Q T. A knowledge element model for knowledge abstract and fusion system[C]// International conference on new trends in information and service science. Beijing: IEEE,2009:23－26.

[53] 谢庆球,秦春秀,杨志娟,等.知识元层次结构表示方法研究[J].情报理论与实践,2017,40(4):26－31.

[54] 王忠义,夏立新,李玉海.基于知识内容的数字图书馆跨学科多粒度知识表示模型构建[J].中国图书馆学报,2019,45(6):50－64.

[55] 索传军,盖双双,周志超.认知计算——单篇学术论文评价的新视角[J].中国图书馆学报,2018,44(1):50－61.

作者贡献说明:

索传军:提出研究选题,设计研究框架,指导论文撰写与修改论文;

戎军涛:调研与整理文献、数据,撰写与修订论文。

Review on the Theory of Knowledge Element

Suo Chuanjun Rong Juntao
School of Information Resource Management, Renmin University of China, Beijing 100872

Abstract: [Purpose/significance] Knowledge element, as the basic component of knowledge, is the existence form of knowledge in microcosmic domain. The research on the composition of knowledge reflects the perception of knowledge, which is the inevitable trend of human knowledge. Therefore, it is of great theoretical value to study the knowledge element and its movement law. [Method/process] Firstly, from the perspective of the discipline history, the research status of the basic components of knowledge and knowledge element theory were combed and analyzed. Furthermore, the existing problems in the research of knowledge element theory were summarized. Finally, the future trend of knowledge element theory was discussed. [Result/conclusion] Some problems can be concluded in the research of knowledge element theory. First of all, the definition of knowledge element is fuzzy. Furthermore, there is ambiguity in the classification of knowledge element types. Thirdly, it is a still difficult task to represent the model of knowledge element. Finally, the theory has not yet gone deep into the study of the essence, structure and movement law of knowledge. We believe that future research involves multiple interdisciplinary perspectives, mainly including the perspective of epistemology combined with philosophy and cognitive science, the perspective of knowledge discovery and mining based on the combination of linguistics and computer science, the perspective of knowledge representation based on the combination of knowledge engineering and library science, and the perspective of knowledge organization research based on the combination of information science and publishing discipline.

Keywords: basic components of knowledge knowledge element types of knowledge element structure of knowledge element knowledge description model